

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-264053

(43)Date of publication of application : 18.09.2002

(51)Int.Cl.

B25J 9/10
 B25J 5/00
 B25J 19/02
 G06T 1/00
 G06T 7/00
 G06T 7/60
 G10L 11/04
 G10L 13/00
 G10L 15/28
 G10L 17/00
 G10L 15/00
 G10L 15/22
 G10L 15/20
 G10L 21/02
 G10L 15/02
 H04N 7/18

(21)Application number : 2001-067849

(22)Date of filing : 09.03.2001

(71)Applicant : JAPAN SCIENCE & TECHNOLOGY CORP

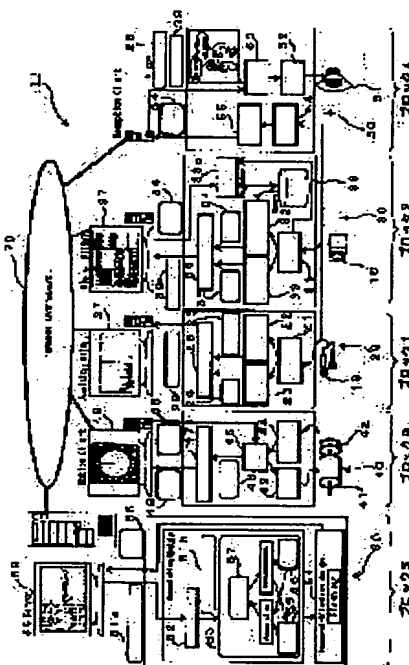
(72)Inventor : NAKADAI KAZUHIRO
 HIDAI KENICHI
 OKUNO HIROSHI
 KITANO HIROAKI

(54) ROBOT AUDIO-VISUAL SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a robot audio-visual system performing the visual and auditory tracing of an object to perform the audio-visual servo of a robot, using both visual and auditory senses.

SOLUTION: An auditory module 20 extracts an auditory event 28 by identifying the sound source of a speaker by pitch extraction and the separation and orientation of the sound source from an acoustic signal of a microphone. A visual module 30 extracts a visual event 39 by the face identification and orientation of the speaker from the image of a camera. A motor control module 40 extracts a motor event 49 from the rotating position of a drive motor. An association module 60 creates an auditory stream 65 and a visual stream 66 from the auditory event, visual event and motor event and creates an association stream 67 by associating these streams. An attention module 64 performs attention control on the basis of the association stream and performs the audio-visual servo of the robot.



LEGAL STATUS

[Date of request for examination]

11.07.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-264053

(P2002-264053A)

(43) 公開日 平成14年9月18日 (2002.9.18)

(51)Int.Cl. ⁷		識別記号	F I		テーマコード(参考)	
B 2 5 J	9/10		B 2 5 J	9/10	A	3 C 0 0 7
	5/00			5/00	C	5 B 0 5 7
	19/02			19/02		5 C 0 5 4
G 0 6 T	1/00	3 4 0	G 0 6 T	1/00	3 4 0 A	5 D 0 1 5
	7/00	3 0 0		7/00	3 0 0 F	5 D 0 4 5
審査請求 未請求 請求項の数4 O L (全 17 頁) 最終頁に続く						

審査請求 未請求 請求項の数 4 O L (全 17 頁) 最終頁に続く

(21) 出願番号 特願2001-67849 (P2001-67849)

(22) 出願日 平成13年3月9日 (2001.3.9)

(71) 出願人 396020800

科学技術振興事業団

埼玉県川口市本町4丁目1番8号

(72) 発明者 中島 一博

千葉県佐倉市白井86

(72) 発明者 日台 健一

埼玉県上尾市上野299-1

(72) 発明者 奥乃 博

東京都渋谷区西原2-10-9

(72) 発明者 北野 宏明

埼玉県川越市西小仙波町2-18-3

(74) 代理人 100082876

弁理士 平山 一幸 (外1名)

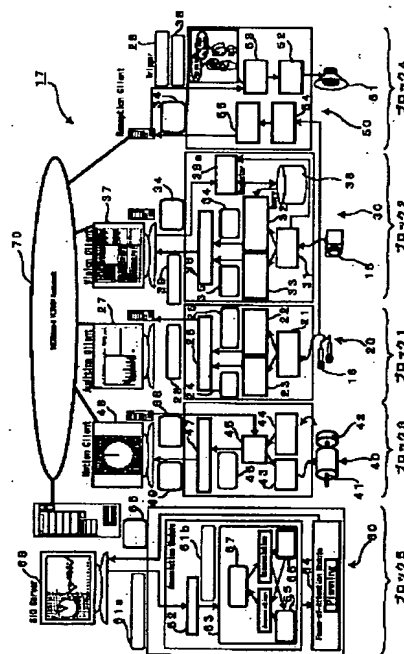
最終頁に続く

(54) 【発明の名称】 ロボット視聴覚システム

(57) 【要約】

【課題】 対象に対する視覚及び聴覚の追跡を行なって、視覚及び聴覚の双方を使用してロボットの視聴覚サーボを行なうようにした、ロボット視聴覚システムを提供する。

【解決手段】 聴覚モジュール20がマイクの音響信号からピッチ抽出、音源の分離及び定位により話者の音源を同定して聴覚イベント28を抽出し、視覚モジュール30がカメラの画像から話者の顔識別と定位により視覚イベント39を抽出し、モータ制御モジュール40が駆動モータの回転位置からモータイベント49を抽出し、アソシエーションモジュール60が聴覚イベント、視覚イベント及びモータイベントから聴覚ストリーム65及び視覚ストリーム66を生成し、これらを関連付けてアソシエーションストリーム67を生成して、アテンション制御モジュール64が、アソシエーションストリームに基づいてアテンション制御を行なって、ロボットの視聴覚サーボを行なう。



【特許請求の範囲】

【請求項1】 外部の音を集音する少なくとも一対のマイクを含む聴覚モジュールと、

ロボットの前方を撮像するカメラを含む視覚モジュールと、

ロボットを水平方向に回転させる駆動モータを含むモータ制御モジュールと、

前記聴覚モジュール、視覚モジュール及びモータ制御モジュールからのイベントを統合してストリームを生成するアソシエーションモジュールと、

アソシエーションモジュールにより生成されたストリームに基づいてアテンション制御を行なうアテンション制御モジュールと、を備えているロボット視聴覚システムであって、

前記聴覚モジュールが、マイクからの音響信号に基づいて、ピッチ抽出、音源の分離及び定位から、少なくとも一人の話者の音源を同定してその聴覚イベントを抽出し、

前記視覚モジュールが、カメラにより撮像された画像に基づいて、各話者の顔識別と定位からその視覚イベントを抽出し、

前記モータ制御モジュールが、駆動モータの回転位置に基づいて、モータイベントを抽出することにより、

前記アソシエーションモジュールが、聴覚イベント、視覚イベント及びモータイベントから、聴覚ストリーム及び視覚ストリームと、これらに関連付けたアソシエーションストリームを生成して、

前記アテンション制御モジュールが、アソシエーションストリームに基づいてモータ制御モジュールの駆動モータ制御のプランニングのためのアテンション制御を行なうこと、ロボットの視聴覚サーボを行なうことを特徴とする、ロボット視聴覚システム。

【請求項2】 外部の音を集音する少なくとも一対のマイクを含む聴覚モジュールと、

ロボットの前方を撮像するカメラを含む視覚モジュールと、

ロボットを水平方向に回転させる駆動モータを含むモータ制御モジュールと、

前記聴覚モジュール、視覚モジュール及びモータ制御モジュールからのイベントを統合してストリームを生成するアソシエーションモジュールと、

アソシエーションモジュールにより生成されたストリームに基づいてアテンション制御を行なうアテンション制御モジュールと、を備えている人型または動物型のロボットの視聴覚システムであって、

前記聴覚モジュールが、マイクからの音響信号に基づいて、ピッチ抽出、音源の分離及び定位から、少なくとも一人の話者の音源を同定してその聴覚イベントを抽出し、

前記視覚モジュールが、カメラにより撮像された画像に

基づいて、各話者の顔識別と定位からその視覚イベントを抽出し、

前記モータ制御モジュールが、駆動モータの回転位置に基づいてモータイベントを抽出することにより、

前記アソシエーションモジュールが、聴覚イベント、視覚イベント及びモータイベントから、聴覚ストリーム及び視覚ストリームと、これらに関連付けたアソシエーションストリームを生成して、

前記アテンション制御モジュールが、アソシエーションストリームに基づいてモータ制御モジュールの駆動モータ制御のプランニングのためのアテンション制御を行なうこと、ロボットの視聴覚サーボを行なうことを特徴とする、ロボット視聴覚システム。

【請求項3】 前記アテンション制御モジュールが、アテンション制御を行なう際に、アソシエーションストリーム、聴覚ストリーム及び視覚ストリームの順に優先させることを特徴とする、請求項1又は2に記載のロボット視聴覚システム。

【請求項4】 前記アテンション制御モジュールが、複数の聴覚ストリーム又は視覚ストリームが存在するとき、状況に応じて一つの聴覚ストリームまたは視覚ストリームを選択し、必要に応じてアソシエーションストリームを生成し、これらの聴覚ストリーム、視覚ストリーム又はアソシエーションストリームに基づいてアテンション制御を行なうことを特徴とする、請求項1から3の何れかに記載のロボット視聴覚システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明はロボット、特に人型または動物型ロボットにおける視聴覚システムに関するものである。

【0002】

【従来の技術】近年、このような人型または動物型ロボットにおいては、視覚、聴覚の能動知覚が注目されている。能動知覚とは、ロボット視覚やロボット聴覚等の知覚を担当する知覚装置を、知覚すべき対象に追従するように、これらの知覚装置を支持する例えば頭部を駆動機構により姿勢制御するものである。

【0003】ここで、能動視覚に関しては、少なくとも知覚装置であるカメラが、駆動機構による姿勢制御によってその光軸方向が対象に向かって保持され、対象に対して自動的にフォーカシングやズームイン、ズームアウト等が行なわれることにより対象がカメラによって撮像されるようになっており、種々の研究が行なわれている。

【0004】これに対して、能動聴覚に関しては、少なくとも知覚装置であるマイクが駆動機構による姿勢制御によって、その指向性が対象に向かって保持され、対象からの音がマイクによって集音される。このような能動聴覚は、例えば本出願人による特願2000-2267

7号(ロボット聴覚システム)に開示されており、視覚情報を参照して音源の方向付けを行なうようにしている。

【0005】

【発明が解決しようとする課題】ところで、これらの能動視覚及び能動聴覚は、ロボットの向き(水平方向)を変更するためのモータ制御モジュールと密接に関連があり、特定の対象に対して能動視覚及び能動聴覚を働かせるためには、ロボットを特定の対象に向ける、即ちアテンション制御を行なう必要がある。しかしながら、従

来、所謂視覚サーボまたは聴覚サーボによるモータモジュールの駆動モータのアテンション制御は行なわれているが、視覚及び聴覚の双方を使用してロボットを正確に制御する、視聴覚サーボは行なわれていない。

【0006】

この発明は、以上の点にかんがみて、対象に対する視覚及び聴覚の追跡を行なって、視覚及び聴覚の双方を使用してロボットの視聴覚サーボを行なうようにした、ロボット視聴覚システムを提供することを目的としている。

【0007】

【課題を解決するための手段】前記目的は、この発明によれば、外部の音を集音する少なくとも一対のマイクを含む聴覚モジュールと、ロボットの前方を撮像するカメラを含む視覚モジュールと、ロボットを水平方向に回転させる駆動モータを含むモータ制御モジュールと、聴覚モジュール、視覚モジュール及びモータ制御モジュールからのイベントを統合してストリームを生成するアソシエーションモジュールと、アソシエーションモジュールにより生成されたストリームに基づいてアテンション制御を行なうアテンション制御モジュールと、を備えているロボット視聴覚システムであって、聴覚モジュールが、マイクからの音響信号に基づいて、ピッチ抽出、音源の分離及び定位から、少なくとも一人の話者の音源を同定してその聴覚イベントを抽出し、視覚モジュールが、カメラにより撮像された画像に基づいて、各話者の顔識別と定位からその視覚イベントを抽出し、モータ制御モジュールが、駆動モータの回転位置に基づいて、モータイベントを抽出することにより、アソシエーションモジュールが、聴覚イベント、視覚イベント及びモータイベントから、聴覚ストリーム及び視覚ストリームと、これらに関連付けたアソシエーションストリームを生成して、アテンション制御モジュールが、アソシエーションストリームに基づいてモータ制御モジュールの駆動モータ制御のプランニングのためのアテンション制御を行なって、ロボットの視聴覚サーボを行なうことを特徴とするロボット視聴覚システムにより、達成される。

【0008】

また、前記目的は、この発明によれば、外部の音を集音する少なくとも一対のマイクを含む聴覚モジュールと、ロボットの前方を撮像するカメラを含む視

モータを含むモータ制御モジュールと、聴覚モジュール、視覚モジュール及びモータ制御モジュールからのイベントを統合してストリームを生成するアソシエーションモジュールと、アソシエーションモジュールにより生成されたストリームに基づいてアテンション制御を行なうアテンション制御モジュールと、を備えている人型または動物型のロボットの視聴覚システムであって、聴覚モジュールが、マイクからの音響信号に基づいて、ピッチ抽出、音源の分離及び定位から少なくとも一人の話者の音源を同定してその聴覚イベントを抽出し、視覚モジュールが、カメラにより撮像された画像に基づいて、各話者の顔識別と定位からその視覚イベントを抽出し、モータ制御モジュールが、駆動モータの回転位置に基づいて、モータイベントを抽出することにより、アソシエーションモジュールが、聴覚イベント、視覚イベント及びモータイベントから、聴覚ストリーム及び視覚ストリームと、これらに関連付けたアソシエーションストリームを生成して、アテンション制御モジュールが、アソシエーションストリームに基づいてモータ制御モジュールの駆動モータ制御のプランニングのためのアテンション制御を行なって、ロボットの視聴覚サーボを行なうことを特徴とするロボット視聴覚システムにより、達成される。

【0009】本発明によるロボット視聴覚システムは、好ましくは、前記アテンション制御モジュールが、アテンション制御を行なう際に、アソシエーションストリーム、聴覚ストリーム及び視覚ストリームの順に優先させる。

【0010】本発明によるロボット視聴覚システムは、好ましくは、前記アテンション制御モジュールが、複数の聴覚ストリーム又は視覚ストリームが存在するとき、状況に応じて一つの聴覚ストリームまたは視覚ストリームを選択し、必要に応じてアソシエーションストリームを生成し、これらの聴覚ストリーム、視覚ストリームまたはアソシエーションストリームに基づいてアテンション制御を行なう。

【0011】前記構成によれば、聴覚モジュールが、マイクが集音した外部の対象からの音から調波構造を利用してピッチ抽出を行なうことにより音源毎の方向を得て、個々の話者の音源を同定し、その聴覚イベントを抽出する。また、視覚モジュールが、カメラにより撮像された画像から、パターン認識による各話者の顔識別と定位から個々の話者の視覚イベントを抽出する。さらに、モータ制御モジュールが、ロボットを水平方向に回転させる駆動モータの回転位置に基づいて、ロボットの方向を検出することによってモータイベントを抽出する。なお、前記イベントとは、各時点において検出される音または顔が検出され、ピッチ及び方向等の特徴が抽出され、話者同定や顔識別等が行なわれること、あるいは駆動モータが回転される状態を示しており、ストリームと

は、時間的に連続するイベントを示している。

【0012】ここで、アソシエーションモジュールは、このようにしてそれぞれ抽出された聴覚イベント、視覚イベント及びモータイベントに基づいて、各話者の聴覚ストリーム及び視覚ストリームを生成し、さらにこれらのストリームを関連付けてアソシエーションストリームを生成して、前記アテンション制御モジュールが、アソシエーションストリームに基づいてアテンション制御を行なうことにより、モータ制御モジュールの駆動モータ制御のプランニングを行なう。アテンションとは、ロボットが対象である話者を、聴覚的及び／又は視覚的に「注目」することであり、アテンション制御とは、モータ制御モジュールによりその向きを変えることにより、ロボットが前記話者に注目するようにすることである。そして、アテンション制御モジュールは、このプランニングに基づいて、モータ制御モジュールの駆動モータを制御することにより、視聴覚サーボによってロボットの方向を対象である話者に向ける。これにより、ロボットが対象である話者に対して正対することにより、聴覚モジュールが当該話者の声を感度の高い正面方向にてマイクにより正確に集音、定位することができる共に、視覚モジュールが当該話者の画像をカメラにより良好に撮像することができるようになる。

【0013】ここで、前記アテンション制御モジュールが、アソシエーションストリームに基づいてアテンション制御を行なうことにより、聴覚情報及び視覚情報の双方を使用して、ロボットの視聴覚サーボを行なうことにより、同一物体からの音声と顔（画像）が同一人に由来していることに基づいて、アテンション制御を行なうことができるので、何れかの情報、即ち聴覚情報又は視覚情報の何れかによる聴覚サーボ又は視覚サーボの場合と比較して、より正確にロボットのサーボを行うことができる。

【0014】従って、このような聴覚モジュール、視覚モジュール及びモータ制御モジュールと、アソシエーションモジュール及びアテンション制御モジュールとの連携によって、ロボットの視聴覚サーボを行なうことにより、ロボットの聴覚及び視覚がそれぞれ有する曖昧性が互いに補完されることになり、所謂ロバスト性が向上し、複数の話者であっても、各話者をそれぞれ知覚することができる。

【0015】

【発明の実施の形態】以下、図面に示した実施形態に基づいて、この発明を詳細に説明する。図1乃至図4はこの発明によるロボット視聴覚システムの一実施形態を備えた実験用の人型ロボットの全体構成を示している。図1において、人型ロボット10は、4DOF（自由度）のロボットとして構成されており、ベース11と、ベース11上に一軸（垂直軸）周りに回転可能に支持された胴体部12と、胴体部12上に、三軸方向（垂直

軸、左右方向の水平軸及び前後方向の水平軸）の周りに揺動可能に支持された頭部13と、を含んでいる。

【0016】前記ベース11は固定配置されていてもよく、脚部として動作可能としてもよい。また、ベース11は移動可能な台車等の上に載置されていてもよい。前記胴体部12は、ベース11に対して垂直軸の周りに、図1にて矢印Aで示すように回転可能に支持されており、図示しない駆動手段によって回転駆動されると共に、図示の場合、防音性の外装によって覆われている。

【0017】前記頭部13は胴体部12に対して連結部材13aを介して支持されており、この連結部材13aに対して前後方向の水平軸の周りに、図1にて矢印Bで示すように揺動可能に、また左右方向の水平軸の周りに、図2にて矢印Cで示すように揺動可能に支持されていると共に、前記連結部材13aが、胴体部12に対してさらに前後方向の水平軸の周りに、図1にて矢印Dで示すように揺動可能に支持されており、それぞれ図示しない駆動手段によって、各矢印A、B、C、D方向に回転駆動される。

【0018】ここで、前記頭部13は、図3に示すように全体が防音性の外装14により覆われていると共に、前側にロボット視覚を担当する視覚装置としてのカメラ15を、また両側にロボット聴覚を担当する聴覚装置としての一对のマイク16（16a、16b）を備えている。

【0019】前記外装14は、例えばウレタン樹脂等の吸音性の合成樹脂から構成されており、頭部13の内部をほぼ完全に密閉することにより、頭部13の内部の遮音を行なうように構成されている。尚、胴体部12の外装も、同様にして吸音性の合成樹脂から構成されている。前記カメラ15は公知の構成であって、例えば所謂パン、チルト、ズームの3DOF（自由度）を有する市販のカメラが適用され得る。

【0020】前記マイク16は、それぞれ頭部13の側面において、前方に向かって指向性を有するように取り付けられている。ここで、マイク16の左右の各マイク16a、16bは、それぞれ図1及び図2に示すように、外装14の両側にて前方に向けた段部14a、14bにて、内側に取り付けられ、段部14a、14bに設けられた貫通穴を通して、前方の音を集音すると共に、外装14の内部の音を拾わないように適宜の手段により遮音されている。これにより、マイク16a、16bは、所謂バイノーラルマイクとして構成されている。なお、マイク16a、16bの取付位置の近傍において、外装14は人間の外耳形状に形成されていてもよい。

【0021】図4は、前記マイク16及びカメラ15を含むロボット視聴覚システムの電氣的構成を示している。図4において、視聴覚システム17は、パーティ受付及びコンパニオン用ロボットとしての構成であり、聴覚モジュール20、視覚モジュール30、モータ制御モ

ジュール40、対話モジュール50、アソシエーションモジュール60及びアテンション制御モジュール64と、から構成されている。以下、図4の各部を拡大して示す図5～図9をも参照しつつさらに説明する。説明の便宜上、聴覚モジュール20をブロック1として図5に拡大して示し、視覚モジュール30をブロック2として図6に拡大して示し、モータ制御モジュール40をブロック3として図7に拡大して示し、対話モジュール50をブロック4として図8に拡大して示し、さらに、アソシエーションモジュール60及びアテンション制御モジュール64をブロック5として図9に拡大して示す。ここで、アソシエーションモジュール60（ブロック5、図9）はサーバから構成されていると共に、他のモジュール、即ち聴覚モジュール20（ブロック1、図5）、視覚モジュール30（ブロック2、図6）、モータ制御モジュール40（ブロック3、図7）、対話モジュール50（ブロック4、図8）は、それぞれクライアントから構成されており、互いに非同期で動作する。

【0022】なお、前記サーバ及び各クライアントは、例えばパーソナルコンピュータにより構成されており、例えば100Base-T等のネットワーク70を介して、例えばTCP/IPプロトコルにより、相互にLAN接続されている。また、各モジュール20、30、40、50、60は、それぞれ階層的に分散して、具体的には下位から順次にデバイス層、プロセス層、特徴層、イベント層から構成されている。

【0023】図5に示すように、前記聴覚モジュール20は、デバイス層としてのマイク16と、プロセス層としてのピーク抽出部21、音源定位部22、音源分離部23と、特徴層（データ）としてのピッチ24、水平方向25と、イベント層としての聴覚イベント生成部26及びビューア27と、から構成されている。

【0024】これにより、聴覚モジュール20は、マイク16からの音響信号に基づいて、ピーク抽出部21により左右のチャンネル毎に一連のピークを抽出して、左右のチャンネルで同じか類似のピークをペアとする。ここで、ピーク抽出は、パワーがしきい値以上で且つ極大値であって、例えば90Hz乃至3kHzの間の周波数であるという条件のデータのみを透過させる帯域フィルタを使用することにより行なわれる。このしきい値は、周囲の暗騒音を計測して自動的に決定される。

【0025】そして、聴覚モジュール20は、各ピーク*

$$BF_{IPD}(\theta) = \int_{-\infty}^{\frac{d(\theta)-m}{\sqrt{\frac{s}{n}}}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

を利用して、IPDの確信度 $BF_{IPD}(\theta)$ を計算する。ここで、 m, s は、それぞれ $d(\theta)$ の平均と分散であり、 n は d の個数である。また、IIDの確信度 $BF_{IID}(\theta)$ は、30度 $<\theta \leq 90$ 度で、前記 I が+の

*が調波構造を有していることを利用して、左右のチャンネル間でより正確なピークのペアを見つけ、左右のチャンネルのピークの各ペアについて、音源分離部23により、逆FFT（高速フーリエ変換）を適用して、各音源からの混合音から調波構造を有する音を分離する。これにより、聴覚モジュール20は、分離した各音について音源定位部22により左右のチャンネルから同じ周波数の音響信号を選択して、例えば5度毎にIPD（両耳間位相差）及びIID（両耳間強度差）を求める。

【0026】そして、聴覚モジュール20の音源定位部22は、所謂聴覚エビポーラ幾何を利用して、ロボット10の正面を0度として ± 90 度の範囲で、仮説推論によるIPD、Phの仮説を生成して、

【数1】

$$d(\theta) = \frac{1}{n_{f < 1.5\text{kHz}}} \sum_{f=0}^{1.5\text{kHz}} \frac{(P_A(\theta, f) - P_r(f))^2}{f}$$

により分離した音と各仮説間の距離 $d(\theta)$ を計算する。ここで、 $n_{f < 1.5\text{kHz}}$ は、周波数が1.5kHz以下である倍音である。これは、左右のマイク15のベースラインからIPDが1.2乃至1.5kHz以下の周波数に対して有効であるので、今回の実験では1.5kHz以下としたものである。

【0027】IIDについては、IPDと同様に、分離した音の各倍音の左右チャンネル間のパワー差から求められる。ただし、IIDについては仮説推論ではなく、

【数2】

$$I = \sum_{f=1.5\text{kHz}}^{3\text{kHz}} I_r(f)$$

による判別関数を用いて、音源が左右何れかを判定するものとする。即ち、周波数 f の各倍音のIIDを $I_r(f)$ としたとき、音源は、 I が正であればロボット10の左方向に、 I が負であれば右方向に、そしてほぼ0であれば正面方向に存在することになる。ここで、IIDの仮説生成には、ロボット10の頭部形状を考慮した膨大な計算が必要となることから、リアルタイム処理を考慮して、IPDと同様の仮説推論は行なわない。

【0028】そして、聴覚モジュール20の音源定位部22は、前記距離 $d(\theta)$ から確立密度関数

【数3】

とき0.35、-のとき0.65、 $-30\text{度} < \theta \leq 30\text{度}$ で、前記 I が+のとき0.5、-のとき0.5、 $-90\text{度} < \theta \leq -30\text{度}$ で、前記 I が+のとき0.65、-のとき0.35となる。

【0029】そして、このようにして得られたIPDの確信度 $BF_{IPD}(\theta)$ 及びIIDの確信度 $BF_{IID}(\theta)$ を、

【数4】

$$BF_{IPD+IID}(\theta) = BF_{IPD}(\theta)BF_{IID}(\theta) + (1 - BF_{IPD}(\theta))BF_{IID}(\theta) + BF_{IPD}(\theta)(1 - BF_{IID}(\theta))$$

で示されるDempster-Shafer理論により統合して、確信度 $BF_{IPD+IID}(\theta)$ を生成する。これにより、聴覚モジュール20は、聴覚イベント生成部26により、音源方向として尤度の高い順に上位20個の確信度 $BF_{IPD+IID}(\theta)$ と方向 (θ) のリストと、ピッチにより、聴覚イベント28を生成する。

【0030】このようにして、聴覚モジュール20は、マイク16からの音響信号に基づいて、ピッチ抽出、音源の分離及び定位から、少なくとも一人の話者の音源を同定してその聴覚イベントを抽出し、ネットワーク70を介してアソシエーションモジュール60に対して送信するようになっている。尚、聴覚モジュール20における上述した処理は、40m秒毎に行なわれる。

【0031】ビューア27は、このようにして生成された聴覚イベント28をクライアントの画面上に表示するものであり、具体的には図11(A)に示すように、左側のウインドウ27aに、聴覚イベント28のパワースペクトルと抽出したピークを、右側のウインドウ27bに、縦軸を相対的な方位角、横軸をピッチ(周波数)とする聴覚イベント28のグラフを表示する。ここで、聴覚イベントは、音源定位の確信度を円の直径とする円により表現されている。

【0032】図6に示すように、前記視覚モジュール30は、デバイス層としてのカメラ15と、プロセス層としての顔発見部31、顔識別部32、顔定位部33と、特徴層(データ)としての顔ID34、顔方向35と、イベント層としての視覚イベント生成部36及びビューア37と、から構成されている。

【0033】これにより、視覚モジュール30は、カメラからの画像信号に基づいて、顔発見部31により例えば肌色抽出により各話者の顔を検出し、顔識別部32にて前もって登録されている顔データベース38により検索して、一致した顔があった場合、その顔ID34を決定して当該顔を識別すると共に、顔定位部33により当該顔方向35を決定(定位)する。なお、顔識別部32による顔データベース38の検索の結果、一致した顔がなかった場合には、顔学習部38aが、顔発見部31が検出した顔を顔データベース38に登録する。

【0034】ここで、視覚モジュール30は、顔発見部31が画像信号から複数の顔を見つけた場合、各顔について前記処理、即ち識別及び定位そして追跡を行なう。その際、顔発見部31により検出された顔の大きさ、方向及び明るさがしばしば変化するので、顔発見部31

は、顔領域検出を行なって、肌色抽出と相関演算に基づくパターンマッチングの組合せによって、200m秒以内に複数の顔を正確に検出できるようになっている。

【0035】また、顔識別部32は、顔発見部31により検出された各顔領域画像を、判別空間に射影し、顔データベース38に前もって登録された顔データとの距離 d を計算する。この距離 d は、登録顔数 (L) に依存するので、

【数5】

$$P_v = \int_{\frac{d^2}{4}}^{\infty} e^{-t} t^{\frac{L}{2}-1} dt$$

により、パラメータの依存しない確信度 P_v に変換される。ここで、判別空間の基底となる判別行列は、公知のオンラインLDAにより、通常のLDAと比較して少ない計算により更新され得るので、リアルタイムに顔データを登録することが可能である。

【0036】顔定位部33は、二次元の画像平面における顔位置を三次元空間に変換し、顔が画像平面にて (x, y) に位置する幅と高さがそれぞれ X 及び Y である $w \times w$ ピクセルであるとする、三次元空間における顔位置は、以下の各式で与えられる方位角 θ 、高さ ϕ 及び距離 r のセットとして得られる。

【数6】

$$r = \frac{C_1}{w}$$

【数7】

$$\theta = \sin^{-1} \left(\frac{x - \frac{X}{2}}{C_2 r} \right)$$

【数8】

$$\phi = \sin^{-1} \left(\frac{\frac{Y}{2} - y}{C_2 r} \right)$$

ここで、 C_1 及び C_2 は、探索画像サイズ (X, Y) とカメラの画角そして実際の顔の大きさにより定義される定数である。

【0037】そして、視覚モジュール30は、各顔毎に、顔ID(名前)34及び顔方向35から、視覚イベント生成部36により視覚イベント39を生成する。詳細には、視覚イベント39は、各顔毎に、上位5つの確信度付きの顔ID(名前)34と位置(距離 r 、水平角度 θ 及び垂直角度 ϕ)から構成されている。

【0038】なお、ビューア37は、視覚イベントをクライアントの画面上に表示するものであり、具体的には図11(B)に示すように、カメラ15による画像37aと、顔識別の確信度付きで抽出した顔の顔IDと定位の結果である位置のリスト37bを表示する。ここで、

カメラ15による画像には、発見し同定された顔が長方形の枠37cで囲まれて表示されている。複数の顔が発見された場合には、各顔について、それぞれ同定を示す長方形の枠37cと、定位の結果としてのリスト37bが表示される。

【0039】図7に示すように、前記モータ制御モジュール40は、デバイス層としてのモータ41及びポテンシオメータ42と、プロセス層としてのPWM制御回路43、AD変換回路44及びモータ制御部45と、特徴層としてのロボット方向46と、イベント層としてのモータイベント生成部47と、ビューア48と、から構成されている。

【0040】これにより、モータ制御モジュール40は、アテンション制御モジュール64（後述）からの指令に基づいてモータ制御部45によりPWM制御回路43を介してモータ41を駆動制御すると共に、モータ41の回転位置をポテンシオメータ42により検出して、AD変換回路44を介してモータ制御部45によりロボット方向46を抽出し、モータイベント生成部47によりモータ方向情報から成るモータイベント49を生成する。

【0041】ビューア48は、モータイベントをクライアントの画面上に三次元的に表示するものであって、具体的には図12（A）に示すように、モータイベント49によるロボット10の向きと動作速度を、例えばOpenGLにより実装されている三次元ビューアを利用してリアルタイムに三次元表示するようになっている。

【0042】図8に示すように、前記対話モジュール50は、デバイス層としてのスピーカ51及びマイク16と、プロセス層としての音声合成回路52、対話制御回路53及び自声抑制回路54、音声認識回路55と、から構成されている。

【0043】これにより、対話モジュール50は、後述するアソシエーションモジュール60により対話制御回路53を制御し、音声合成回路52によりスピーカ51を駆動することによって、対象とする話者に対して所定の音声を発すると共に、マイク16からの音響信号から自声抑制回路54によりスピーカ51からの音を除去した後、音声認識回路55により対象とする話者の音声を認識する。なお、前記対話モジュール50は、階層としての特徴層及びイベント層を備えていない。

【0044】ここで、対話制御回路53は、例えばパーティ受付ロボットの場合には、現在のアテンションを継続することが最優先となるが、パーティロボットの場合には、最も最近に関連付けられたストリームに対して、アテンション制御される。

【0045】図9に示すように、前記アソシエーションモジュール60は、上述した聴覚モジュール20、視覚モジュール30、モータ制御モジュール40、対話モジュール50に対して、階層的に上位に位置付けられてお

り、各モジュール20、30、40、50のイベント層の上位であるストリーム層を構成している。具体的には、前記アソシエーションモジュール60は、聴覚モジュール20、視覚モジュール30及びモータ制御モジュール40からの非同期イベント61a即ち聴覚イベント28、視覚イベント39及びモータイベント49を同期させて同期イベント61bにする同期回路62と、これらの同期イベント61bを相互に関連付けて、聴覚ストリーム65、視覚ストリーム66及びアソシエーションストリーム67を生成するストリーム生成部63と、さらにアテンション制御モジュール64と、ビューア68を備えている。

【0046】前記同期回路62は、聴覚モジュール20からの聴覚イベント28、視覚モジュール30からの視覚イベント38及びモータ制御モジュール40からのモータイベント49を同期させて、同期聴覚イベント、同期視覚イベント及び同期モータイベントを生成する。その際、聴覚イベント28及び視覚イベント38は、同期モータイベントによって、その座標系が絶対座標系に変換されることになる。

【0047】ここで、各イベントの実際に観測されてからネットワーク70を介してアソシエーションモジュール60に到着するまでの遅延時間は、例えば聴覚イベント28では4.0m秒、視覚イベント39では200m秒、モータイベント49では100m秒であり、ネットワーク70における遅延が10乃至200m秒であり、さらに到着周期も異なることによるものである。従って、各イベントの同期を取るために、聴覚モジュール20、視覚モジュール30及びモータ制御モジュール40からの聴覚イベント28、視覚イベント39及びモータイベント49は、それぞれ実際の観測時間を示すタイムスタンプ情報を備えており、図示しない短期記憶回路にて、例えば2秒間の間だけ一旦記憶される。

【0048】そして、同期回路62は、短期記憶回路に記憶された各イベントを、上述した遅延時間を考慮して、実際の観測時間と比較して500m秒の遅延時間を備えるように、同期プロセスにより取り出す。これにより、同期回路62の応答時間は500m秒となる。また、このような同期プロセスは例えば100m秒周期で動作するようになっている。尚、各イベントは、それぞれ互いに非同期でアソシエーションモジュール60に到着するので、同期を取るための観測時刻と同時刻のイベントが存在するとは限らない。従って、同期プロセスは、同期を取るための観測時刻前後に発生したイベントに対して、線形補間による補間を行なうようになっている。

【0049】また、ストリーム生成部63は、以下の点に基づいて、ストリーム65、66、67の生成を行なう。

1. 聴覚イベント28は、同等または倍音関係にある

ピッチを備え、方向が±10度以内で最も近い聴覚ストリーム65に接続される。なお、±10度以内の値は、聴覚エピソード幾何の精度を考慮して選定されたものである。

2. 視覚イベント39は、共通の顔ID34を有し且つ40cmの範囲内で最も近い視覚ストリーム66に接続される。なお、40cmの範囲内の値は、秒速4m以上で人間が移動することがないということを前提として選定されたものである。

3. すべてのストリームに対して探索を行なった結果、接続可能なストリーム65、66が存在しないイベントがある場合には、当該イベント28、39は、新たなストリーム65、66を構成することになる。

4. 既に存在しているストリーム65、66は、これらに接続されるイベント28、39がない場合には、最大で500m秒間は存続するが、その後もイベントが接続されない状態が継続すると、消滅する。

5. 聴覚ストリーム65と視覚ストリーム66が±10度以内に近接する状態が1秒間のうち500m秒以上継続する場合、これの聴覚ストリーム65と視覚ストリーム66は、同一話者に由来するものとみなされ、互いに関係付けられて、アソシエーションストリーム67が生成される。

6. アソシエーションストリーム67は、聴覚イベント28または視覚イベント39が3秒間以上接続されない場合には、関係付けが解除され、既存の聴覚ストリーム65または視覚ストリーム66のみが存続する。

7. アソシエーションストリーム67は、聴覚ストリーム65及び視覚ストリーム66の方向差が3秒間、±30度以上になった場合には、関係付けが解除され、個々の聴覚ストリーム65及び視覚ストリーム66に戻る。

【0050】これにより、ストリーム生成部63は、同期回路62からの同期聴覚イベント及び同期視覚イベントに基づいて、これらの時間的つながりを考慮してイベントを接続することにより、聴覚ストリーム65及び視覚ストリーム66を生成すると共に、相互の結び付きの強い聴覚ストリーム65及び視覚ストリーム66を関係付けて、アソシエーションストリーム67を生成するようになっており、逆にアソシエーションストリーム67を構成する聴覚ストリーム65及び視覚ストリーム66の結び付きが弱くなれば、関係付けを解除するようになっている。

【0051】また、アテンション制御モジュール64は、モータ制御モジュール40の駆動モータ制御のプランニングのためのアテンション制御を行なうものであり、その際アソシエーションストリーム67、聴覚ストリーム65そして視覚ストリーム66の順に優先的に参照して、アテンション制御を行なう。そして、アテンション制御モジュール64は、聴覚ストリーム65及び視

覚ストリーム66の状態とアソシエーションストリーム67の存否に基づいて、ロボット10の動作プランニングを行ない、駆動モータ41の動作の必要があれば、モータ制御モジュール40に対して動作指令としてのモータイベントをネットワーク70を介して送信する。

【0052】ここで、アテンション制御モジュール64におけるアテンション制御は、連続性とトリガに基づいており、連続性により同じ状態を保持しようとし、トリガにより最も興味のある対象を追跡しようとする。

従って、アテンション制御は、1. アソシエーションストリームの存在は、ロボット10に対して正対して話している人が現在も存在している、あるいは近い過去に存在していたことを示している、このようなロボット10に対して話している人に対して、高い優先度でアテンションを向けて、トラッキングを行なう必要がある。

2. マイク16は無指向性であるので、カメラの視野角のような検出範囲が存在せず、広範囲の聴覚ストリームを得ることができるので、視覚ストリームより聴覚ストリームの優先度を高くすべきである。という二つの点を考慮して、以下の原則に従ってアテンションを向けるストリームを選択して、トラッキングを行なう。

1. アソシエーションストリームのトラッキングを最優先する。

2. アソシエーションストリームが存在しない場合、聴覚ストリームのトラッキングを優先する。

3. アソシエーションストリーム及び聴覚ストリームが存在しない場合、視覚ストリームのトラッキングを優先する。

4. 複数の同じ種類のストリームが存在する場合、最も古いストリームのトラッキングを優先する。

このようにして、アテンション制御モジュール64は、聴覚情報及び視覚情報に基づいて生成されたアソシエーションストリームによりアテンション制御を行なって、ロボットの視聴覚サーボによりモータ制御モジュール40の駆動モータ41の制御のプランニングを行ない、このプランニングに基づいてモータコマンド66を生成し、ネットワーク70を介してモータ制御モジュール40に伝送する。これにより、モータ制御モジュール40では、このモータコマンド66に基づいてモータ制御部45がPWM制御を行なって、駆動モータ41を回転駆動させて、ロボット10を所定方向に向けるようになっている。

【0053】ビューア68は、このようにして生成された各ストリームをサーバの画面上に表示するものであり、具体的には図12(B)に示すように、レーダチャート68a及びストリームチャート68bにより表示する。ここで、レーダチャート68aは、その瞬間におけるストリームの状態、より詳細には広く明るい扇形で示されるカメラの視野角68a1と、狭く暗い扇形で示さ

れる音源方向68a2を示す。また、ストリームチャート68bは、太線により示されるアソシエーションストリーム68b1と、細線により示される聴覚ストリームまたは視覚ストリーム68b2を示している。

【0054】本発明実施形態による人型ロボット10は以上のように構成されており、パーティ受付ロボットとして対象とする話者に対して、図10を参照して、以下のように動作する。まず、図10(A)に示すように、ロボット10は、パーティ会場の入口前に配置されている。そして、図10(B)に示すように、パーティ参加者Pがロボット10に接近してくるが、ロボット10は、まだ当該参加者Pを認識していない。ここで、参加者Pがロボット10に対して例えば「こんにちは」と話し掛けると、ロボット10は、マイク16が当該参加者Pの音声を拾って、聴覚モジュール20が音源方向を伴う聴覚イベント28を生成して、ネットワーク70を介してアソシエーションモジュール60に伝送する。

【0055】これにより、アソシエーションモジュール60は、この聴覚イベント28に基づいて聴覚ストリーム29を生成する。このとき、視覚モジュール30は参加者Pがカメラ15の視野内に入っていないので、視覚イベント39を生成しない。従って、アソシエーションモジュール60は、聴覚イベント28のみに基づいて聴覚ストリーム29を生成し、アテンション制御モジュール64は、この聴覚ストリーム29をトリガーとして、ロボット10を参加者Pの方向に向けるようなアテンション制御を行なう。

【0056】このようにして、図10(C)に示すように、ロボット10が参加者Pの方向を向き、所謂声によるトラッキングが行なわれる。そして、視覚モジュール30がカメラ15による参加者Pの顔の画像を取り込んで、視覚イベント39を生成して、当該参加者Pの顔を顔データベース38により検索し、顔識別を行なうと共に、その結果である顔ID34及び画像をネットワーク70を介してアソシエーションモジュール60に伝送する。なお、当該参加者Pの顔が顔データベース38に登録されていない場合には、視覚モジュール30は、その旨をネットワーク70を介してアソシエーションモジュールに伝送する。

【0057】このとき、ロボット10は、聴覚イベント28及び視覚イベント39によりアソシエーションストリーム65を生成しており、このアソシエーションストリーム65により視聴覚サーボを行なうことにより、アテンション制御モジュール64は、そのアテンション制御を変更しないので、ロボット10は、参加者Pの方向を向き続ける。従って、参加者Pが移動したとしても、ロボット10は、アソシエーションストリーム65によりモータ制御モジュール40を制御することにより参加者Pを追跡して、視覚モジュール30のカメラ15が参加者Pを継続して撮像し得るようになっている。

【0058】そして、アソシエーションモジュール60は、聴覚モジュール20の音声認識回路55に入力を与えて、音声認識回路55はその音声認識結果を対話制御回路53に与える。これにより、対話制御回路53は、音声合成を行なってスピーカ51から発声する。このとき、音声認識回路55がマイク16からの音響信号からスピーカ51からの音を自声抑制回路54により低減することにより、ロボット10は自身の発声を無視して相手の声をより正確に認識することができる。

【0059】ここで、音声合成による発声は、参加者Pの顔が前記顔データベース38に登録されているか否かで異なる。参加者Pの顔が顔データベース38に登録されている場合には、アソシエーションモジュール60は、視覚モジュール30からの顔ID34に基づいて、対話モジュール50を制御して、音声合成により「こんにちは。XXXさんですか？」と参加者Pに対して質問する。これに対して、参加者Pが「はい。」と答えると、対話モジュール50がマイク16からの音響信号に基づいて、音声認識回路55により「はい」を認識して、対話制御回路53により音声合成を行ない、スピーカ51から「ようこそXXXさん、どうぞ部屋にお入り下さい。」と発声する。

【0060】また、参加者Pの顔が顔データベース38に登録されていない場合には、アソシエーションモジュール60は、対話モジュール50を制御して、音声合成により「こんにちは。あなたのお名前を教えてくださいませんか？」と参加者Pに対して質問する。これに対して、参加者Pが「XXXです。」と自分の名前を答えると、対話モジュール50がマイク16からの音響信号に基づいて、音声認識回路55により「XXX」を認識して、対話制御回路53により音声合成を行ない、スピーカ51から「ようこそXXXさん、どうぞ部屋にお入り下さい。」と発声する。このようにして、ロボット10は、参加者Pの認識を行なって、図10(D)に示すように、パーティ会場への入場を誘導すると共に、視覚モジュール30にて、当該参加者Pの顔の画像と名前「XXX」を顔データベース38に登録させる。

【0061】また、人型ロボット10は、コンパニオン用ロボットとして、図13及び図14を参照して、以下のように動作する。まず、人型ロボット10は、特に明確なシナリオを有しているのではなく、例えば図13に示すシナリオをベンチマークとして使用して、本システムの評価を行なった。なお、図14(A)はロボット方向、図14(B)は視覚ストリームによるトラッキング、図14(C)は聴覚ストリームによるトラッキングを示している。このシナリオでは、二人の話者A、Bが約40秒間に亘って種々のアクションを行なう。前記シナリオは、具体的には以下の通りである。

時刻t1：A氏がロボット10の視野内に入る。視覚モジュール30がA氏の顔を検出して視覚イベントを生成

し、アソシエーションモジュール60により視覚ストリームが生成される。

時刻t2: A氏がロボット10に対して話し始める。聴覚モジュール20がA氏の声を検出して聴覚イベントを生成し、アソシエーションモジュール60により聴覚ストリーム65が生成され、さらにアソシエーションストリーム67が生成される。これにより、ロボットの視聴覚サーボが行なわれる。

時刻t3: B氏がロボット10の視野外で話し始める。聴覚モジュール20が(見えない)B氏の声を検出して、聴覚イベントを生成し、アソシエーションモジュール60により聴覚ストリームが生成される。

時刻t4: A氏が移動して、物陰に隠れる。これにより、視覚モジュール30がA氏の視覚イベントを生成しなくなり、A氏の視覚ストリームが途切れるが、アソシエーションストリームは所定時間の間継続する。

時刻t5: A氏が再び物陰から現われる。これにより、視覚モジュール30がA氏の視覚イベントを再び生成し、アソシエーションモジュール60により、再びアソシエーションストリーム67が生成される。その後、A氏は話を止めて、再び物陰に隠れる。視覚モジュール30がA氏の視覚イベントを生成しなくなり、A氏の視覚ストリームが途切れるので、所定時間後にアソシエーションが解除され、アソシエーションストリーム67が消滅する。

時刻t7: 聴覚ストリームをトリガーとして、ロボット10が話をしているB氏の方を向く。

時刻t8: ロボット10がB氏を視野内に捉える。視覚モジュール30がB氏の視覚イベントを生成し、アソシエーションモジュール60によりB氏の視覚ストリームが生成され、さらにB氏のアソシエーションストリーム67が生成される。

時刻t9: A氏が話をしながら、ロボット10の視野内に入ってくる。聴覚モジュール20及び視覚モジュール30がA氏の聴覚イベント及び視覚イベントを生成し、アソシエーションモジュール60がA氏の聴覚ストリーム及び視覚ストリームが生成される。

時刻t10: B氏が話を止める。聴覚モジュール20がB氏の聴覚イベントを生成しなくなり、アソシエーションモジュール60がB氏のアソシエーションを解除してB氏の聴覚ストリームは消滅し、視覚ストリームのみが残る。そして、ロボット10がアテンションをA氏に向けると共に、同様にしてA氏のアソシエーションストリーム67が生成される。

【0062】このようにして、上述したシナリオにおいては、本発明によるロボット視聴覚システムにおいては、以下のような特徴を有することが分かる。

1. 時刻t1及びt6にて、新しいアソシエーションストリームが生成されると、アテンション制御モジュール64におけるアテンションが新しいアソシエーション

に向けられる。

2. 時刻t4、t5にて、A氏が見えなくなることにより、アソシエーションストリームの視覚ストリームが欠落したときであっても、アソシエーションが継続していることにより、聴覚ストリームによるA氏のトラッキングが継続され得る。

3. 時刻t6、t11にて、アソシエーションストリームが消滅することにより、アソシエーションストリームの次に優先度の高い聴覚ストリームによりアテンション制御が行なわれ、話者のトラッキングが行なわれ、図13に示すように、ロボット10がトラッキングの対象である話者に正対して、当該話者からの音声をマイク15の正面方向により確実に検出することができるようになってい

4. 時刻t9以降、二人の話者A氏及びB氏は、同時にカメラ15の視野内に収まる程度(方向差約20度)に接近しているが、この場合でも、二人の聴覚ストリーム、視覚ストリーム及びアソシエーションストリームは、それぞれ明確に別個に生成され、各話者のトラッキングが確実に行なわれる。

【0063】このようにして、人型ロボット10は、聴覚イベント28及び視覚イベント39が生成される場合には、これらを互いに関連付けて、アソシエーションストリーム67を生成して、このアソシエーションストリーム67に基づいてアテンション制御を行なうことにより、ロボットの視聴覚サーボを行なうことになる。従って、従来の聴覚サーボまたは視覚サーボの場合と比較して、聴覚及び視覚の双方を使用することによって、話者をより正確に追跡することが可能になる。また、途中で話者が物陰に隠れたり視野外に移動して見えなくなると、図14(B)(視覚イベントの第一候補のみを示す)に示すように視覚ストリームによるトラッキングが途切れた場合には、図14(C)に示すように、聴覚ストリームによるアソシエーションストリーム67によって、当該話者を確実にトラッキングすることかできるので、常に複数の話者を聴覚及び視覚により認識していると共に、複数の話者のうちの一人の話者を追跡したり、あるいは途中で他の話者に切り換えて追跡することができる。

【0064】なお、図14(B)において、時刻t4及びt5の間では視覚ストリームが途切れ、また時刻t6からt7までの間は話者がロボット10の視野外に居ることから、視覚ストリームに基づいて、話者のトラッキングを行なうことはできないが、図14(C)に示す聴覚ストリームを参照することによって、話者のトラッキングを確実に行なうことができる。また、図14(C)において、時刻t3が23秒付近まで、そして34秒付近から時刻t10の間は、正しくA氏及びB氏の二本の聴覚ストリームが分離されているが、時刻t8及びt6の周辺では、誤った聴覚ストリームが生成されている。

また、時刻t5から17秒付近までの間は、A氏の移動及びロボット11の水平回転が同時に行なわれているため、話者の移動及びモータノイズそしてそのエコーにより音源からの音響信号の品質が低下しており、二人の話者の定位はあまり正確ではない。このような場合でも、図14(B)に示す視覚ストリームを参照することにより、話者のトラッキングを確実にこなうことができる。このようにして、聴覚ストリーム及び視覚ストリームが互いに関連付けられてアソシエーションストリームが生成される場合には、聴覚及び視覚の双方を使用して、ロボットの視聴覚サーボを行なうことによって、聴覚ストリーム及び視覚ストリームがそれぞれ有する曖昧性が互いに補完されることにより、所謂ロバスト性が向上し、複数の話者であっても、各話者をそれぞれ確実に知覚して、トラッキングを行なうことができる。

【0065】また、コンパニオン用ロボットとしての人型ロボット10は、パーティ受付ロボットと顔データベース38を共用し、あるいはパーティ受付ロボットの顔データベース38が転送または複写されるようにしてもよい。この場合、コンパニオン用ロボットとしての人型ロボット10は、パーティ参加者全員を常に顔識別によって認識することができる。

【0066】このようにして、本発明実施形態による人型ロボット10によれば、聴覚モジュール20及び視覚モジュール30からの聴覚イベント及び視覚イベントに基づいて、アソシエーションモジュール60が聴覚ストリーム、視覚ストリームそしてアソシエーションストリームを生成することによって、複数の対象である話者を視聴覚により認識しているので、聴覚または視覚のいずれか一方のみによるサーボの場合と比較して、より正確に話者の追跡を行なうことができると共に、何れかのイベントが欠落したり明確に認識できなくなった場合には、例えば話者が移動して「見えなく」なった場合でも聴覚により、また話者が話をせず「聞こえなく」なった場合でも視覚により、リアルタイムに複数の話者を聴覚的及び／又は視覚的にトラッキングすることができる。

【0067】上述した実施形態において、人型ロボット10は、4DOF(自由度)を有するように構成されているが、これに限らず任意の動作を行なうように構成されたロボットに本発明によるロボット聴覚システムを組み込むことも可能である。また、上述した実施形態においては、本発明によるロボット視聴覚システムを人型ロボット10に組み込んだ場合について説明したが、これに限らず、大型等の各種動物型ロボットや、その他の形式のロボットに組み込むことも可能であることは明らかである。

【0068】

【発明の効果】以上述べたように、この発明によれば、聴覚モジュール、視覚モジュール及びモータ制御モジュールと、アソシエーションモジュール及びアテンション

制御モジュールとの連携によって、聴覚及び視覚の双方を使用して、ロボットの視聴覚サーボを行なうことにより、ロボットの聴覚及び視覚がそれぞれ有する曖昧性が互いに補完されることになり、所謂ロバスト性が向上し、複数の話者であっても各話者をそれぞれ知覚することができる。また、例えば聴覚イベントまたは視覚イベントの何れか一方が欠落したときであっても、視覚イベントまたは聴覚イベントのみに基づいて、対象である話者をアソシエーションモジュールが知覚することができるので、リアルタイムにモータ制御モジュールの制御を行なうことができる。さらに、聴覚ストリーム及び視覚ストリームのうち、状況に応じて、双方または一方のみを利用して、話者のトラッキングを行なうことにより、常により一層正確な話者のトラッキングを行なうことができると共に、同時に複数の聴覚ストリーム及び視覚ストリームが存在していても、これらの聴覚ストリーム及び視覚ストリームに基づいて、そのときの状況に応じて、これらの聴覚ストリーム及び視覚ストリームの何れかを適宜に利用することにより、各話者のトラッキングをより確実に行なうことができる。これにより、本発明によれば、対象に対する視覚及び聴覚の追跡を行なって、視覚及び聴覚の双方を使用してロボットの視聴覚サーボを行なうようにした、極めて優れたロボット視聴覚システムが提供される。

【図面の簡単な説明】

【図1】この発明によるロボット聴覚装置の第一の実施形態を組み込んだ人型ロボットの外観を示す正面図である。

【図2】図1の人型ロボットの側面図である。

【図3】図1の人型ロボットにおける頭部の構成を示す概略拡大図である。

【図4】図1の人型ロボットにおけるロボット視聴覚システムの電氣的構成を示すブロック図である。

【図5】図4におけるブロック1の聴覚モジュールを拡大して示す電氣的構成のブロック図である。

【図6】図4におけるブロック2の視覚モジュールを拡大して示す電氣的構成のブロック図である。

【図7】図4におけるブロック3のモータ制御モジュールを拡大して示す電氣的構成のブロック図である。

【図8】図4におけるブロック4の対話モジュールを拡大して示す電氣的構成のブロック図である。

【図9】図4におけるブロック5のアソシエーションモジュールを拡大して示す電氣的構成のブロック図である。

【図10】図4のロボット視聴覚システムにおけるパーティ受付ロボットとしての動作例を示す図である。

【図11】図4のロボット視聴覚システムにおける(A)聴覚モジュール、(B)視覚モジュールのビューアの画面例を示す図である。

【図12】図4のロボット視聴覚システムにおける

21

22

(A) モータ制御モジュール、(B) アソシエーションモジュールのビューアの画面例を示す図である。

【図13】図4のロボット視聴覚システムにおけるコンパニオン用ロボットとしての動作例を示す各時刻における(A)レーダチャート、(B)ストリームチャート及び(C)カメラ画像を示す図である。

【図14】図13の動作例における(A)ロボット方向、(B)視覚ストリーム及び(C)聴覚ストリームを示すグラフである。

【符号の説明】

10 人型ロボット

11 ベース

12 胴体部

* 13 頭部

13 a 連結部材

14 外装

15 カメラ(ロボット視覚)

16, 16 a, 16 b マイク(ロボット聴覚)

17 ロボット視聴覚システム

20 聴覚モジュール

30 視覚モジュール

40 モータ制御モジュール

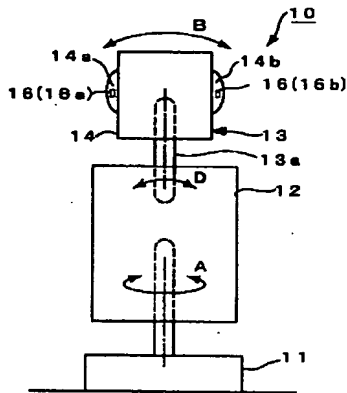
10 50 対話モジュール

60 アソシエーションモジュール

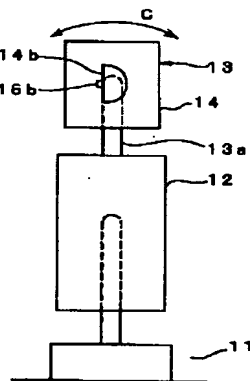
64 アテンション制御モジュール

* 70 ネットワーク

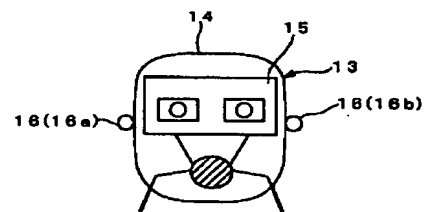
【図1】



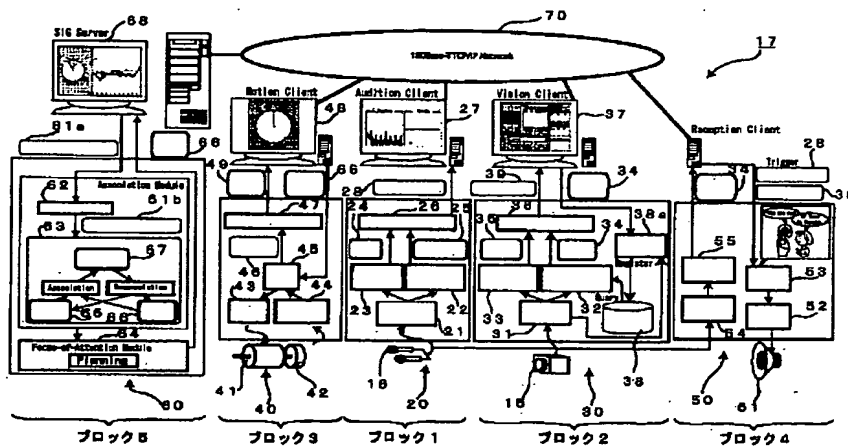
【図2】



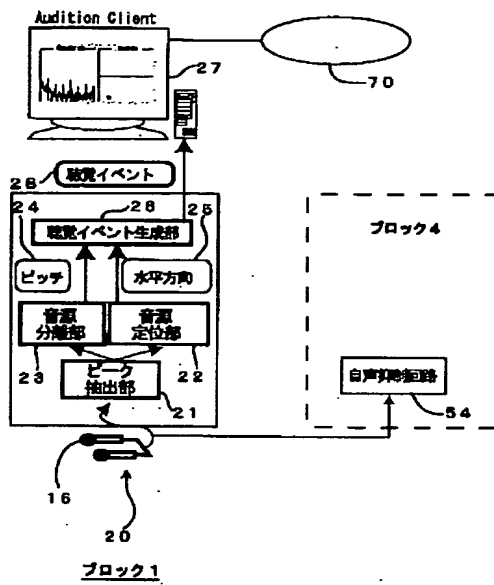
【図3】



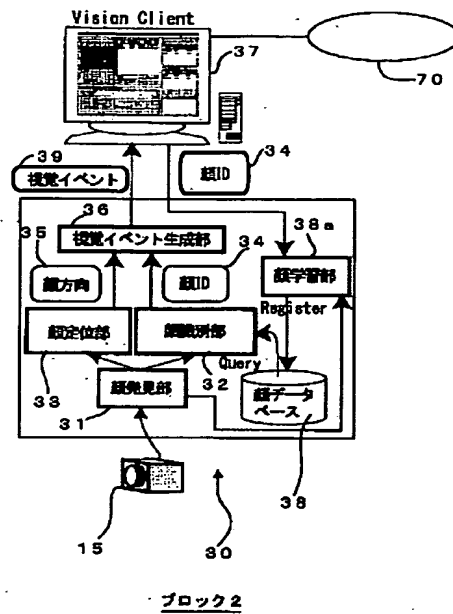
【図4】



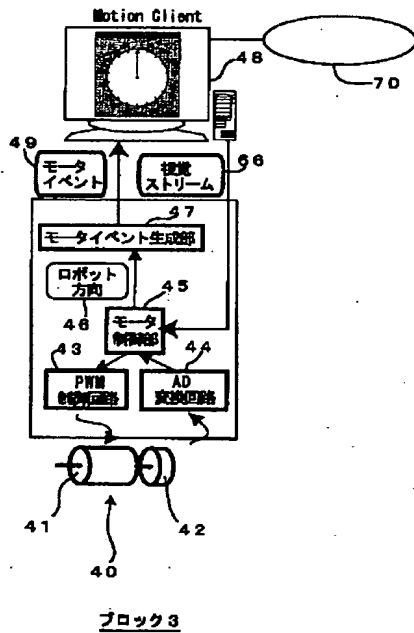
【図5】



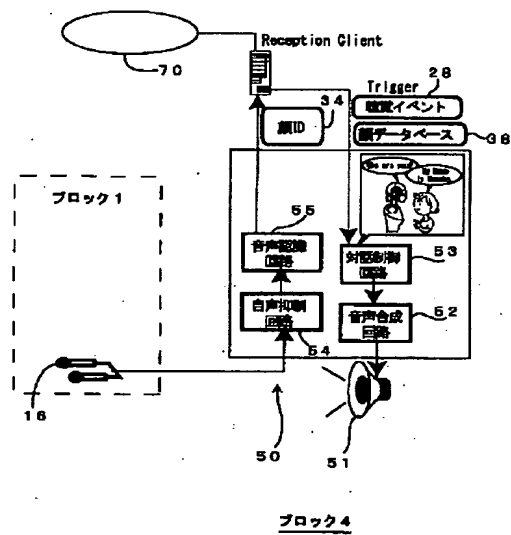
【図6】



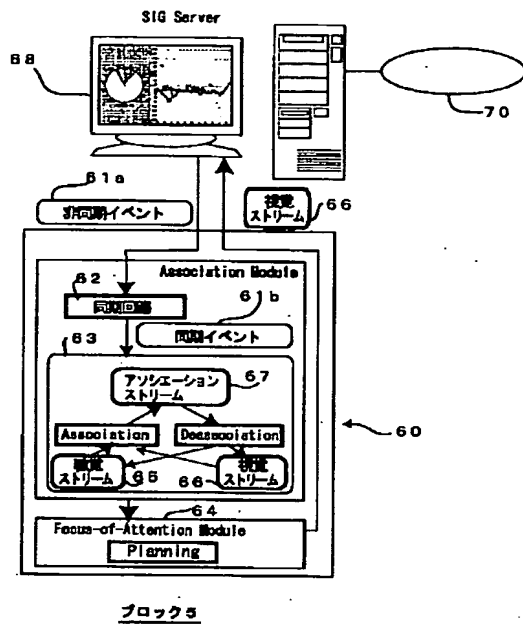
【図7】



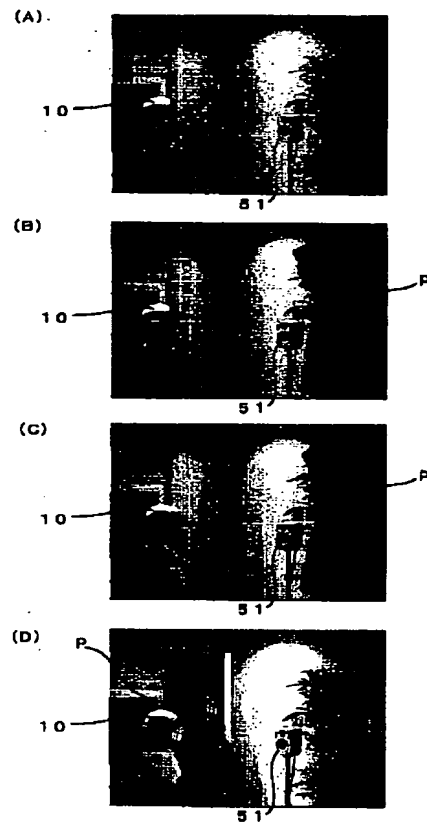
【図8】



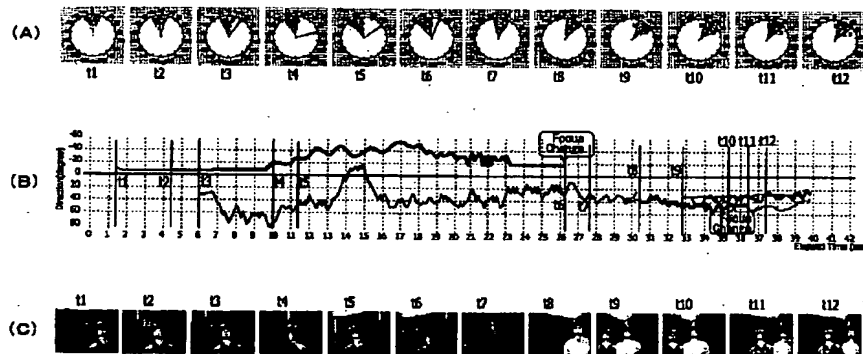
【図9】



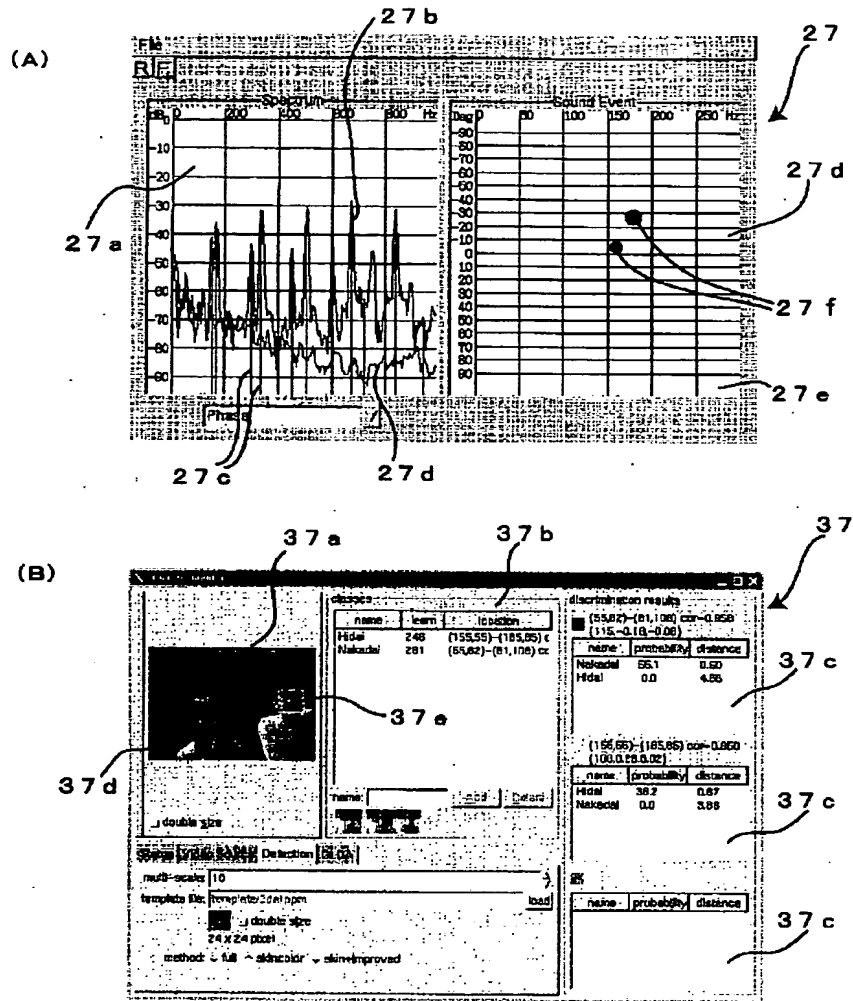
【図10】



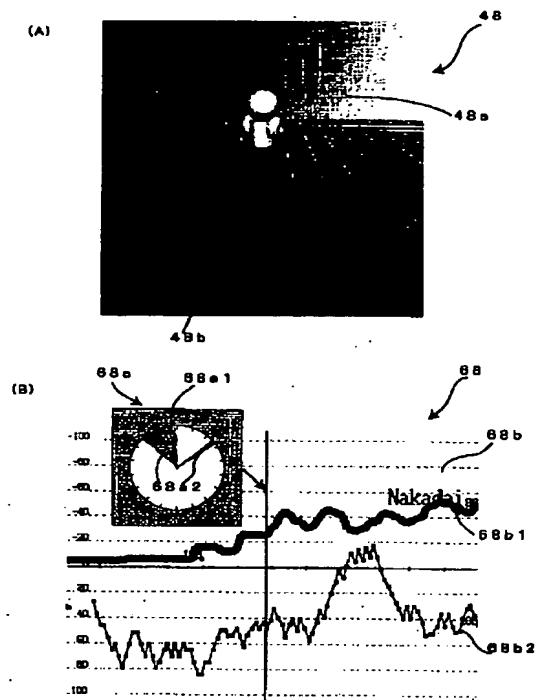
【図13】



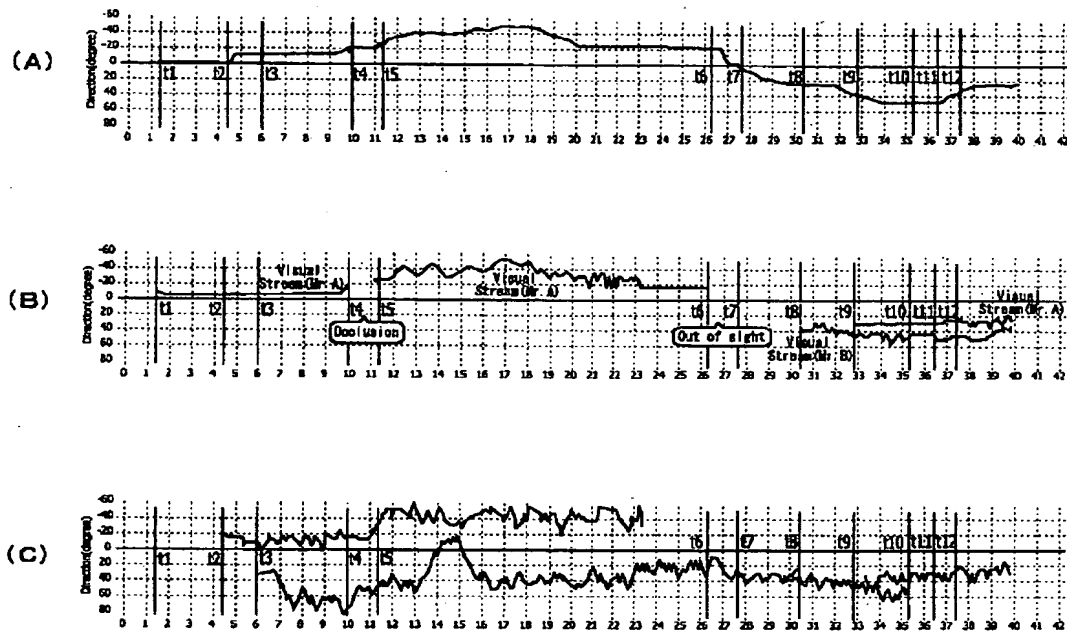
【図11】



【図12】



【図14】



フロントページの続き

(51)Int.Cl.	識別記号	F I	ターム (参考)
G 0 6 T 7/60	1 5 0	G 0 6 T 7/60	1 5 0 B 5 L 0 9 6
G 1 0 L 11/04		H 0 4 N 7/18	Z
13/00		G 1 0 L 3/00	C
15/28			Q
17/00			5 1 1
15/00			5 4 5 F
15/22			5 5 1 H
15/20			5 7 1 T
21/02		3/02	3 0 1 C
15/02		9/00	3 0 1 A
H 0 4 N 7/18			

F ターム (参考) 3C007 AS34 AS36 CS08 JS03 KS04
 KS08 KS18 KS20 KS39 KT01
 KT11 KT15 LT08 NS01 WA02
 WA03 WB19 WC07 WC16
 5B057 AA05 BA02 CA12 CA16 DA06
 DB02
 5C054 AA01 CA04 CA08 CC05 CD03
 CG06 EF06 FC12 FF07 HA04
 5D015 AA03 CC13 DD02 EE04 KK01
 KK04 LL06
 5D045 AB11
 5L096 BA05 CA02 FA69 HA09